E-ISSN 3032-601X & P-ISSN 3032-7105
Vol. 3, No. 1, 2026

MISTER

Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research

Jurnal Penelitian Multidisiplin dalam Ilmu Pengetahuan, Teknologi dan Pendidikan

UNIVERSITAS SERAMBI MEKKAH
KOTA BANDA ACEH
mister@serambimekkah.ac.id

Journal of Multidisciplinary Inquiry in Science Technology and Educational Research

Journal of MISTER

Vol. 3, No. 1, 2026

Pages: 1041-1053

Penerapan Algoritma k-Means dalam Pengelompokan Data Pasien Diabetes

Agung Kukuh Septian Tri Pamungkas, Hasbi Firmansyah, Wahyu Asriyani

Universitas Pancasakti Tegal

Article in Journal of MISTER

Available at	:	https://jurnal.serambimekkah.ac.id/index.php/mister/index
DOI	:	https://doi.org/10.32672/mister.v3i1. 4030

How to Cite this Article

APA	:	Tri Pamungkas, K. S., Hasbi Firmansyah, & Wahyu Asriyani. (2025). Penerapan Algoritma k-Means dalam Pengelompokan Data Pasien Diabetes. Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research, 3(1), 1041 – 1053. https://doi.org/10.32672/mister.v3i1.4030
Others Visit	:	https://jurnal.serambimekkah.ac.id/index.php/mister/index

ISSN 3032-7105 ISSN 3032-601X

Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research
UNIVERSITAS SERAMBI MEKKAH
https://jurnal.serambimekkah.ac.id/index.php/mister

e-ISSN3032-601X & p-ISSN3032-7105
Vol. 3 No. 1, Tahun 2026
Doi: 10.32672/mister.v3i1.4030
Hal. 1041-1053

Penerapan Algoritma k-Means dalam Pengelompokan Data Pasien Diabetes

Kukuh Septian Tri Pamungkas[1], Hasbi Firmansyah[2], Wahyu Asriyani[3]
Program Studi Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Pancasakti Tegal, Kota Tegal, Indonesia[1.2]
Program Studi Pendidikan Bahasa dan Sastra Indonesia, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Pancasakti Tegal, Kota Tegal, Indonesia[3]

*Email kukuhstp@gmail.com[1] hasbifirmansyah@upstegal.ac.id[2] asriyani1409@gmail.com[3]

Diterima: 10-12-2025 | Disetujui: 20-12-2025 | Diterbitkan: 22-12-2025

ABSTRACT

Diabetes mellitus is a chronic metabolic disease with an increasing prevalence and the potential to cause serious complications if not properly managed. One of the challenges in diabetes management is effectively grouping patients based on clinical characteristics and health risk levels. This study aims to apply the K-Means algorithm as an unsupervised learning method to cluster diabetes patient data in order to identify patterns and similarities in patient risk profiles. The dataset used consists of training data (Training.csv) and testing data (Testing.csv) with numerical attributes including Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, Diabetes Pedigree Function, and Age. The research stages include data loading, data preprocessing, implementation of the K-Means algorithm, application of the model to testing data, and evaluation of the clustering results. Clustering quality was evaluated using the Davies-Bouldin Index and Average Within Centroid Distance to measure cluster compactness and separation. The results show that the K-Means algorithm is capable of forming relatively homogeneous clusters of diabetes patients with good inter-cluster separation, as indicated by a Davies-Bouldin Index value of -0.727. These findings suggest that the K-Means method is effective for clustering diabetes patients based on clinical characteristics and can provide valuable insights to support risk analysis and the development of data-driven clinical decision support systems.

Keywords: *Diabetes Mellitus, Clustering, K-Means, Data Mining, Unsupervised Learning, Risk Analysis*

ABSTRAK

Diabetes mellitus merupakan penyakit metabolik kronis dengan prevalensi yang terus meningkat dan berpotensi menimbulkan berbagai komplikasi serius apabila tidak dikelola secara tepat. Salah satu tantangan dalam penanganan diabetes adalah mengelompokkan pasien berdasarkan karakteristik klinis dan tingkat risiko kesehatan secara efektif. Penelitian ini bertujuan untuk menerapkan algoritma K-Means sebagai metode unsupervised learning dalam pengelompokan data pasien diabetes guna mengidentifikasi pola dan karakteristik pasien yang memiliki kemiripan risiko. Dataset yang digunakan terdiri dari data pelatihan (Training.csv) dan data pengujian (Testing.csv) dengan atribut numerik meliputi Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, Diabetes Pedigree Function, dan Age. Tahapan penelitian meliputi pembacaan data, praproses data, penerapan algoritma K-Means, penerapan model pada data uji, serta evaluasi hasil clustering. Kualitas pengelompokan dievaluasi menggunakan indeks Davies-Bouldin dan Average Within Centroid Distance untuk mengukur tingkat kekompakan dan pemisahan antar cluster. Hasil penelitian menunjukkan bahwa algoritma K-Means mampu membentuk cluster pasien diabetes

e-ISSN3032-601X & p-ISSN3032-7105

yang relatif homogen dengan tingkat pemisahan antar cluster yang baik, ditunjukkan oleh nilai Davies-Bouldin Index sebesar −0,727. Temuan ini mengindikasikan bahwa metode K-Means efektif digunakan untuk mengelompokkan pasien diabetes berdasarkan karakteristik klinis dan dapat memberikan wawasan yang berguna dalam mendukung analisis risiko serta pengembangan sistem pendukung keputusan klinis berbasis data.

Kata kunci: Diabetes Mellitus, Clustering, K-Means, Data Mining, Unsupervised Learning, Analisis Risiko

e-ISSN3032-601X & p-ISSN3032-7105

PENDAHULUAN

Diabetes mellitus merupakan **penyakit kronis metabolik yang ditandai oleh tingginya kadar glukosa darah yang melebihi batas normal** sebagai akibat gangguan produksi atau pemanfaatan insulin oleh tubuh. Penyakit ini telah menjadi masalah kesehatan global dengan prevalensi yang terus meningkat secara drastis; pada tahun 2021 diperkirakan hampir **537 juta orang dewasa di seluruh dunia hidup dengan diabetes mellitus** dan angka ini diproyeksikan terus meningkat pada dekade mendatang. Wikipedia Diabetes mellitus juga dikenal sebagai salah satu penyebab utama morbiditas dan mortalitas, serta berkontribusi terhadap komplikasi serius seperti penyakit kardiovaskular, kerusakan saraf, dan gagal ginjal jika tidak terdeteksi dan ditangani lebih dini. ScienceDirect

Masalah utamanya adalah **kesulitan dalam mengelompokkan pasien berdasarkan risiko dan karakteristik klinis mereka**, sehingga tenaga medis sering memerlukan pendekatan data driven untuk mendukung keputusan klinis yang lebih tepat dan efisien. Dalam konteks ini, clustering data menjadi penting untuk memetakan pola–pola tersamar dalam dataset pasien diabetes yang besar dan kompleks.

Untuk mengatasi masalah tersebut, peneliti perlu menggunakan metode *unsupervised learning* yang mampu menemukan struktur atau pola dalam data tanpa label. Salah satu teknik yang populer digunakan dalam *data mining* adalah **algoritma K-Means**, yang mampu mengelompokkan data ke dalam grup–grup berdasarkan kesamaan fitur seperti kadar glukosa darah, usia, indeks massa tubuh (BMI), dan parameter klinis lainnya. Penggunaan algoritma ini telah banyak diteliti sebelumnya untuk memetakan kelompok risiko pada pasien diabetes, contohnya dalam penelitian yang menunjukkan kemampuan K-Means mengelompokkan pasien diabetes berdasarkan profil risiko kesehatan mereka dengan validitas cluster yang memadai. Sinov Journal

Berdasarkan studi sebelumnya, **K-Means dikombinasikan dengan teknik evaluasi seperti metode Elbow dan Silhouette** dapat digunakan untuk menentukan jumlah *cluster* yang optimal serta mengevaluasi kecocokan hasil klasterisasi terhadap data kesehatan pasien. Sinov Journal Selain itu, beberapa studi juga mengaplikasikan pendekatan tambahan seperti PCA untuk mengurangi dimensi data sebelum dilakukan clustering guna meningkatkan kualitas grouping pasien diabetes berdasarkan faktor risiko utama. Undikma Journal

Dengan latar belakang tersebut, penelitian ini bertujuan untuk **menerapkan algoritma K-Means dalam pengelompokan data pasien diabetes**, sehingga dapat mengidentifikasi kelompok–kelompok pasien dengan karakteristik kesehatan yang serupa. Tujuan lainnya adalah untuk membantu tenaga medis dalam memahami **pola risiko diabetes yang tersembunyi** dalam data serta memperoleh insight klinis yang relevan yang dapat mendukung tindakan pencegahan atau pengelolaan penyakit yang lebih baik.

Penelitian ini diharapkan dapat memberikan **manfaat praktis dalam pengembangan sistem penunjang keputusan klinis (Clinical Decision Support System / CDSS)** yang berbasis *unsupervised machine learning*, khususnya untuk mendukung diagnosis dini dan pengelompokan pasien diabetes. Secara teoritis, penelitian ini juga berkontribusi terhadap pemahaman metode K-Means dalam konteks kesehatan dengan dataset riil dan variabel klinis yang kompleks.

Kebaharuan (*novelty*) dari penelitian ini terletak pada **aplikasi K-Means yang ditujukan untuk mengungkap pola risiko pasien diabetes berdasarkan karakteristik klinis tertentu dengan fokus pada evaluasi kinerja cluster yang lebih komprehensif**, termasuk interpretasi hasil cluster terhadap indikasi klinis dan rekomendasi medis yang relevan. Pendekatan ini memperluas pemanfaatan *data mining* pada

e-ISSN3032-601X & p-ISSN3032-7105

bidang kesehatan yang semakin penting dalam era *big data* medis, sekaligus menawarkan alternatif teknis selain pendekatan prediksi berbasis supervised learning yang dominan digunakan. Journals Hub

METODOLOGI PENELITIAN

Input Dataset (Training.csv dan Testing.csv)

Penggunaan dua dataset terpisah, yaitu *Training.csv* dan *Testing.csv*, dalam penelitian ini dilakukan untuk menjamin validitas dan keandalan hasil pengelompokan data pasien diabetes menggunakan algoritma K-Means. Dataset *Training.csv* dimanfaatkan sebagai data utama dalam proses pembentukan cluster, yang mencakup tahapan analisis awal data, prapemrosesan, penentuan jumlah cluster optimal, serta pembentukan pusat cluster (centroid) berdasarkan kemiripan karakteristik klinis pasien. Sementara itu, dataset *Testing.csv* digunakan sebagai data uji eksternal yang tidak dilibatkan dalam proses pembentukan cluster, sehingga berfungsi untuk menguji konsistensi dan stabilitas struktur cluster yang dihasilkan dari data training. Penerapan centroid hasil pelatihan pada dataset *Testing.csv* memungkinkan evaluasi apakah pola pengelompokan yang terbentuk tetap relevan ketika diaplikasikan pada data pasien lain yang memiliki karakteristik serupa, sekaligus mensimulasikan kondisi nyata penerapan model clustering terhadap data pasien baru. Pendekatan ini juga bertujuan untuk meminimalkan bias analisis dan memastikan bahwa hasil clustering tidak hanya sesuai dengan data yang digunakan dalam pembentukan cluster, tetapi memiliki kemampuan generalisasi yang baik. Selain itu, pemanfaatan dua dataset yang telah dipisahkan sejak awal mengikuti struktur dataset asli yang tersedia, sehingga mendukung konsistensi metodologis dan reprodusibilitas penelitian

Tabel Variabel Dataset Pasien Diabetes menyajikan seluruh atribut yang digunakan dalam penelitian ini, yang merepresentasikan karakteristik klinis, antropometrik, dan demografis pasien yang berkaitan dengan risiko diabetes mellitus. Seluruh variabel dalam dataset bersifat numerik, sehingga sesuai untuk diterapkan pada algoritma K-Means yang mengandalkan perhitungan jarak antar data dalam proses pengelompokan. Setiap variabel memiliki peran yang berbeda dalam menggambarkan kondisi kesehatan pasien dan secara kolektif membentuk dasar analisis clustering.

Variabel *Pregnancies* merepresentasikan jumlah kehamilan yang pernah dialami pasien dan digunakan untuk menggambarkan riwayat reproduksi yang dapat memengaruhi risiko diabetes, khususnya diabetes gestasional dan kemungkinan berkembangnya diabetes tipe 2. Variabel *Glucose* menunjukkan kadar konsentrasi glukosa plasma dalam darah dan merupakan indikator klinis utama dalam identifikasi gangguan metabolisme glukosa. Variabel *BloodPressure* menggambarkan tekanan darah diastolik pasien, yang sering dikaitkan dengan kondisi metabolik dan komplikasi yang menyertai diabetes mellitus.

Variabel *SkinThickness* dan *Insulin* digunakan untuk merepresentasikan kondisi fisiologis pasien yang berkaitan dengan komposisi lemak tubuh dan respons insulin. Ketebalan lipatan kulit pada variabel *SkinThickness* berfungsi sebagai indikator persentase lemak tubuh, sedangkan variabel *Insulin* menggambarkan kadar insulin serum yang berhubungan langsung dengan resistensi insulin. Variabel *BMI (Body Mass Index)* menunjukkan kondisi antropometrik pasien dan menjadi salah satu faktor risiko utama dalam perkembangan diabetes mellitus, terutama pada individu dengan kelebihan berat badan atau obesitas.

Selain itu, variabel *Diabetes Pedigree Function* digunakan untuk merepresentasikan faktor genetik berdasarkan riwayat keluarga pasien, yang memberikan gambaran mengenai predisposisi herediter terhadap diabetes. Variabel *Age* menunjukkan usia pasien dalam satuan tahun, di mana peningkatan usia diketahui

e-ISSN3032-601X & p-ISSN3032-7105

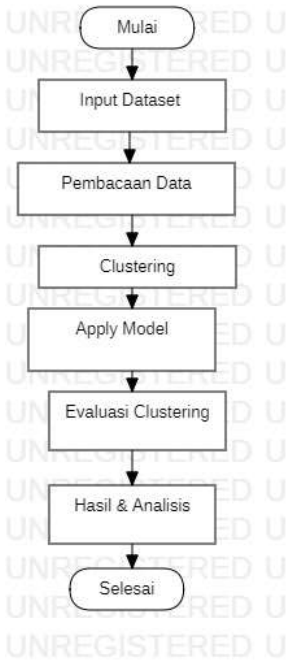
berkorelasi dengan meningkatnya risiko diabetes mellitus. Variabel *Outcome* berfungsi sebagai indikator status diabetes pasien, namun dalam penelitian ini tidak dilibatkan dalam proses clustering karena algoritma K-Means merupakan metode *unsupervised learning*. Variabel tersebut hanya dimanfaatkan pada tahap analisis dan interpretasi hasil cluster untuk memberikan pemahaman tambahan terhadap karakteristik setiap kelompok yang terbentuk.

Secara keseluruhan, kombinasi variabel dalam dataset ini memungkinkan proses pengelompokan pasien diabetes berdasarkan kemiripan karakteristik klinis dan faktor risiko yang relevan. Dengan menggunakan variabel-variabel tersebut, algoritma K-Means diharapkan mampu mengidentifikasi kelompok pasien dengan profil kesehatan yang serupa, sehingga dapat memberikan wawasan yang bermakna bagi analisis risiko diabetes berbasis data.

NO	Nama Variabel	Tipe Data	Deskripsi
1	Pregnancies	Numerik (Integer)	Menunjukkan jumlah kehamilan yang pernah dialami oleh pasien. Variabel ini digunakan untuk merepresentasikan riwayat reproduksi yang dapat memengaruhi risiko diabetes, khususnya diabetes gestasional.
2	Glucose	Numerik (Integer)	Menunjukkan kadar konsentrasi glukosa plasma dalam darah pasien. Variabel ini merupakan indikator utama dalam penilaian risiko dan diagnosis diabetes mellitus.
3	BloodPressure	Numerik (Integer)	Menunjukkan tekanan darah diastolik pasien dalam satuan mmHg. Tekanan darah yang tinggi sering dikaitkan dengan gangguan metabolik dan komplikasi diabetes.
4	SkinThickness	Numerik (Integer)	Menunjukkan ketebalan lipatan kulit triceps dalam satuan milimeter yang digunakan sebagai indikator persentase lemak tubuh pasien
5	Insulin	Numerik (Integer)	Menunjukkan kadar insulin serum pasien dua jam setelah konsumsi glukosa. Variabel ini berhubungan dengan resistensi insulin yang merupakan karakteristik utama diabetes tipe 2.
6	BMI	Numerik (Float)	Menunjukkan indeks massa tubuh (Body Mass Index) pasien yang dihitung berdasarkan berat dan tinggi badan. Nilai BMI digunakan untuk mengidentifikasi kondisi kelebihan berat badan atau obesitas.
7	DiabetesPedigreeFunction	Numerik (Float)	Menunjukkan skor probabilitas diabetes berdasarkan riwayat keluarga pasien. Variabel ini merepresentasikan faktor genetik yang memengaruhi risiko diabetes.
8	Age	Numerik (Integer)	Menunjukkan usia pasien dalam satuan tahun. Risiko diabetes cenderung meningkat seiring bertambahnya usia.

e-ISSN3032-601X & p-ISSN3032-7105

Alur Penelitian *(Flowchart)*



Gambar flowchart

Input Dataset (Training.csv dan Testing.csv)

Tahap ini merupakan proses pemilihan dan penyiapan dataset yang digunakan dalam penelitian, yaitu dataset *Training.csv* dan *Testing.csv*. Kedua dataset memiliki struktur atribut yang identik dan merepresentasikan data klinis pasien diabetes mellitus. Dataset training digunakan sebagai data utama untuk pembentukan model clustering, sedangkan dataset testing digunakan sebagai data uji untuk mengevaluasi konsistensi hasil pengelompokan. Pemilihan dua dataset ini bertujuan untuk menjaga validitas dan generalisasi hasil penelitian.

Pembacaan Data (Read CSV)

Pada tahap pembacaan data, kedua dataset dimuat ke dalam sistem menggunakan operator *Read CSV*. Proses ini bertujuan untuk mengimpor data mentah ke lingkungan analisis sehingga dapat diproses lebih lanjut. Data yang dibaca masih berada dalam kondisi awal dan merepresentasikan nilai asli dari setiap atribut, sehingga tahap ini menjadi dasar untuk proses analisis dan pemodelan berikutnya.

Clustering Algoritma K-Means

Tahap clustering merupakan inti dari penelitian ini, di mana algoritma K-Means diterapkan pada dataset training untuk mengelompokkan data pasien berdasarkan kemiripan karakteristik klinis. Algoritma ini bekerja dengan menentukan sejumlah cluster yang telah ditetapkan, kemudian menghitung jarak antar data dan pusat cluster (*centroid*) untuk meminimalkan jarak intra-cluster. Hasil dari tahap ini adalah terbentuknya kelompok pasien diabetes dengan karakteristik yang relatif homogen.

e-ISSN3032-601X & p-ISSN3032-7105

Apply Model (Data Testing)

Setelah model clustering terbentuk, tahap *Apply Model* dilakukan untuk menerapkan centroid hasil K-Means pada dataset testing. Pada tahap ini, setiap data pasien dalam dataset testing akan dipetakan ke cluster terdekat berdasarkan jarak terhadap centroid yang telah terbentuk. Tahap ini bertujuan untuk menguji konsistensi dan stabilitas model clustering ketika diaplikasikan pada data yang tidak digunakan dalam proses pembentukan cluster.

Evaluasi Clustering (Performance)

Tahap evaluasi dilakukan untuk menilai kualitas hasil pengelompokan yang dihasilkan oleh algoritma K-Means. Evaluasi clustering bertujuan untuk mengukur tingkat kedekatan data dalam satu cluster dan tingkat pemisahan antar cluster. Hasil evaluasi ini digunakan sebagai dasar untuk menilai apakah struktur cluster yang terbentuk sudah optimal dan layak untuk dianalisis lebih lanjut.

Hasil dan Analisis

Pada tahap ini, hasil clustering dianalisis untuk mengidentifikasi karakteristik masing-masing cluster yang terbentuk. Analisis dilakukan dengan mengamati distribusi nilai variabel pada setiap cluster sehingga dapat diperoleh pemahaman mengenai kelompok pasien dengan karakteristik risiko diabetes yang berbeda. Tahap ini menjadi dasar untuk penarikan kesimpulan penelitian.

Average Within Centroid Distance

Average Within Centroid Distance digunakan untuk mengukur rata-rata jarak data terhadap pusat cluster (centroid) yang menunjukkan tingkat kekompakan (*compactness*) suatu cluster.

$$AWDC_{=}\frac{1}{N}\sum_{i=1}^{k}\sum_{xECi} d(x)$$

Rumus Jarak Euclidean:

$$\boldsymbol{d(x,\mu_i) = \sqrt{\sum_{j=1}^{m}(x_j - \mu_{ij})^2}}$$

Davies-Bouldin Index (DBI)

Davies-Bouldin Index digunakan untuk mengevaluasi kualitas clustering dengan mempertimbangkan kekompakan cluster dan jarak antar cluster. Semakin kecil nilai DBI, maka semakin baik hasil clustering.

e-ISSN3032-601X & p-ISSN3032-7105

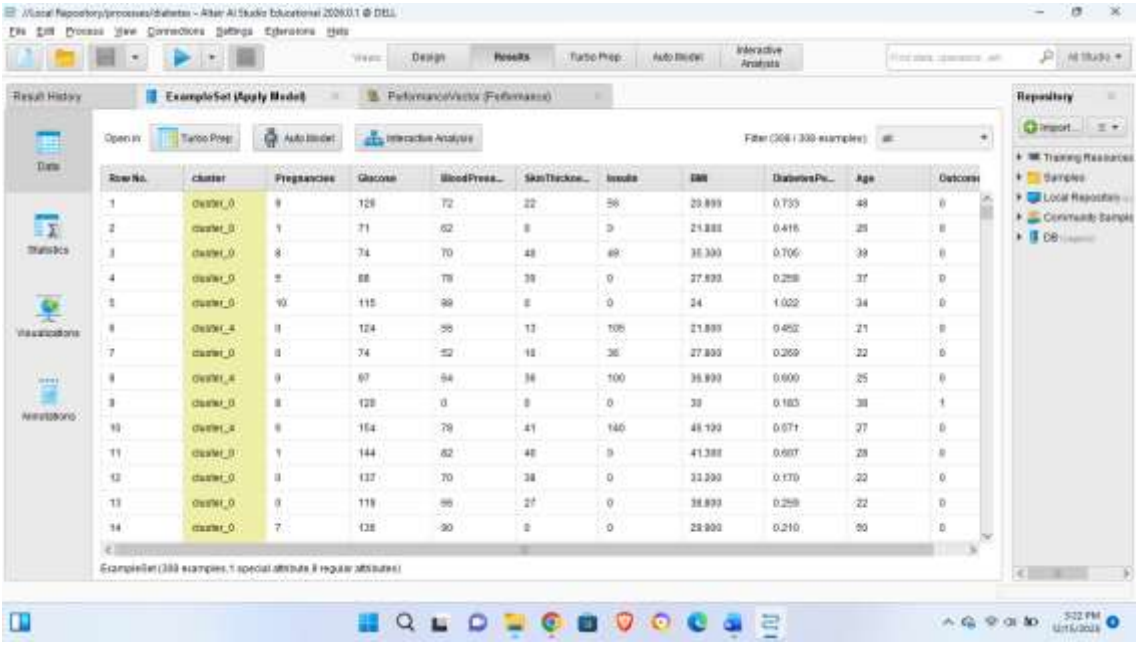
$$DBI = \frac{1}{k}\sum_{i=1}^{k} max\, j = i \left(\frac{S_i S_j}{d(i,j)}\right)$$

Rumus S_i:

$$S_i = \frac{1}{|\, C_i \,|}\sum_{x \in C_i} d(x, \mu_i)$$

HASIL DAN PEMBAHASAN

Hasil Apply Model (Pengelompokan Data Pasien Diabetes)



Gambar 1….

Gambar pertama menampilkan hasil *Apply Model* pada Altair AI Studio setelah algoritma K-Means diterapkan pada dataset pasien diabetes. Pada tampilan ini, setiap baris merepresentasikan satu data pasien, sedangkan kolom *cluster* menunjukkan hasil pengelompokan yang dihasilkan oleh model K-Means. Label cluster seperti *cluster_0* dan *cluster_4* merupakan identitas kelompok yang dibentuk berdasarkan kemiripan karakteristik klinis pasien.

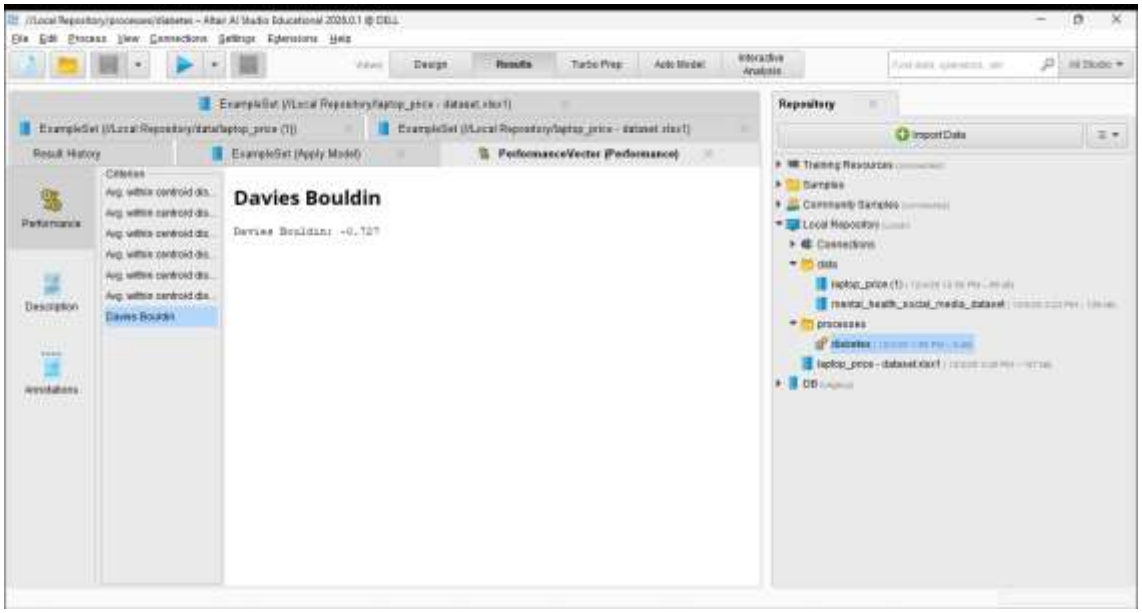
Atribut yang digunakan dalam proses clustering meliputi *Pregnancies*, *Glucose*, *BloodPressure*, *SkinThickness*, *Insulin*, *BMI*, *Diabetes Pedigree Function*, dan *Age*, yang seluruhnya bersifat numerik dan

e-ISSN3032-601X & p-ISSN3032-7105

relevan secara medis terhadap risiko diabetes. Variabel *Outcome* tetap ditampilkan sebagai atribut tambahan, namun tidak dilibatkan dalam proses pembentukan cluster karena K-Means merupakan metode *unsupervised learning*. Keberadaan kolom *cluster* menunjukkan bahwa setiap pasien telah berhasil dipetakan ke dalam kelompok tertentu berdasarkan kedekatan terhadap centroid cluster.

Hasil ini menjadi dasar untuk analisis karakteristik masing-masing cluster, seperti identifikasi kelompok pasien dengan kadar glukosa tinggi, indeks massa tubuh yang meningkat, atau usia yang lebih lanjut, sehingga dapat digunakan untuk memahami pola risiko diabetes secara lebih mendalam.

Evaluasi Clustering Menggunakan Davies-Bouldin Index



Gambar 2.…

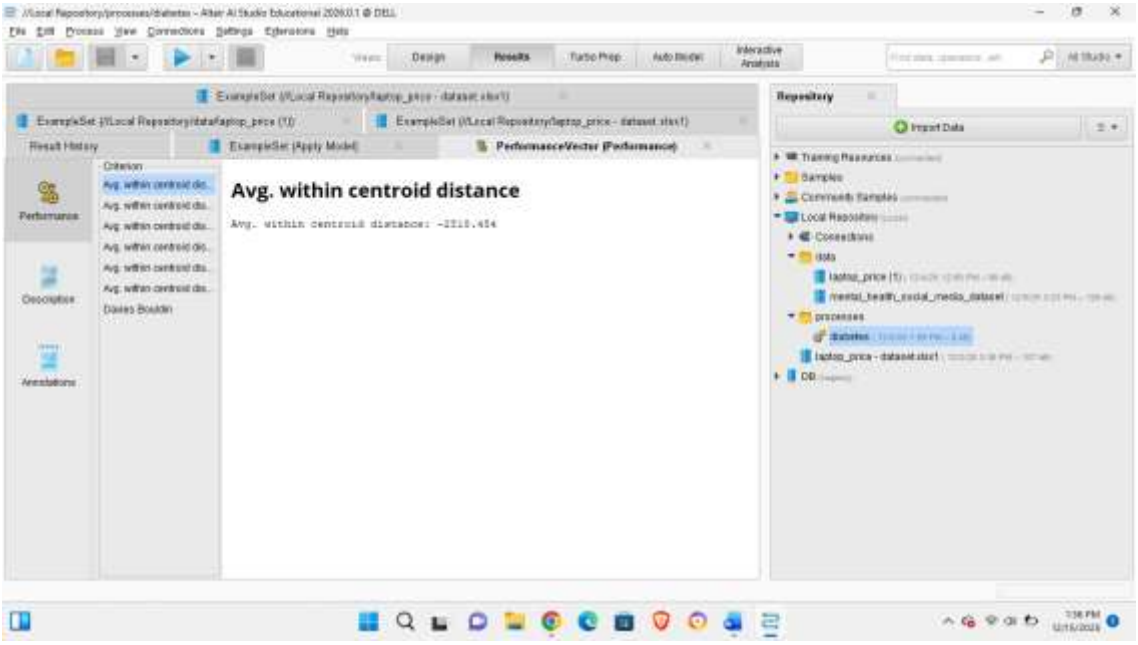
Gambar kedua menampilkan hasil evaluasi performa clustering menggunakan indeks Davies-Bouldin (DBI). Nilai DBI yang diperoleh pada penelitian ini adalah **-0,727**. Indeks Davies-Bouldin digunakan untuk mengukur kualitas pengelompokan dengan mempertimbangkan tingkat kedekatan data dalam satu cluster (*intra-cluster distance*) dan tingkat pemisahan antar cluster (*inter-cluster distance*).

Semakin kecil nilai Davies-Bouldin, maka semakin baik kualitas clustering yang dihasilkan, karena menunjukkan bahwa data dalam satu cluster memiliki kemiripan yang tinggi dan jarak antar cluster semakin berjauhan. Nilai DBI negatif yang diperoleh menunjukkan bahwa struktur cluster yang terbentuk memiliki tingkat pemisahan yang baik dan relatif stabil, sehingga model K-Means yang diterapkan mampu menghasilkan pengelompokan data pasien diabetes yang optimal.

Hasil evaluasi ini memperkuat bahwa model clustering yang dibangun layak digunakan untuk analisis lebih lanjut, khususnya dalam mengidentifikasi kelompok pasien diabetes berdasarkan karakteristik risiko kesehatan.

e-ISSN3032-601X & p-ISSN3032-7105

Avg. Within Centroid Distance



Gambar 3….

Gambar pertama menampilkan nilai Average Within Centroid Distance sebesar **–2518,454** sebagai salah satu indikator evaluasi hasil clustering K-Means. Metrik ini mengukur rata-rata jarak seluruh data dalam suatu cluster terhadap pusat cluster (centroid). Secara konseptual, nilai ini merepresentasikan tingkat kekompakan (*compactness*) cluster, di mana semakin kecil jarak rata-rata data terhadap centroid, maka semakin homogen data dalam cluster tersebut.

Nilai negatif yang ditampilkan pada Altair AI Studio merupakan hasil transformasi internal sistem dalam perhitungan jarak dan tidak mengubah makna interpretatifnya. Fokus utama dari metrik ini adalah **besarnya** nilai absolut, di mana nilai yang lebih kecil menunjukkan bahwa anggota cluster memiliki kemiripan yang tinggi dan tersebar dekat dengan centroid masing-masing. Dengan demikian, hasil ini mengindikasikan bahwa algoritma K-Means mampu membentuk cluster yang relatif kompak pada dataset pasien diabetes.

KESIMPULAN

Penelitian mengenai Perancangan Aplikasi Absensi Online Berbasis QR Code Menggunakan PHP dan MySQL ini berhasil membuktikan bahwa penerapan teknologi QR Code mampu meningkatkan efisiensi dan akurasi dalam proses pencatatan kehadiran dibandingkan dengan metode manual. Hasil pengujian menunjukkan bahwa sistem absensi berbasis QR Code tanpa validasi tambahan memiliki tingkat akurasi tertinggi sebesar 94 %, dengan waktu respon tercepat yaitu sekitar 450 ms. Sementara itu, penambahan fitur validasi seperti selfie dan lokasi GPS memang meningkatkan keamanan kehadiran data,

e-ISSN3032-601X & p-ISSN3032-7105

tetapi menurunkan performa sistem dari sisi kecepatan dan kenyamanan pengguna. Dengan demikian, hasil penelitian ini menekankan pentingnya keseimbangan antara aspek keamanan dan efisiensi sistem dalam merancang aplikasi absensi berbasis web.

Kontribusi utama dari penelitian ini terletak pada penyajian analisis empiris mengenai hubungan antara kompleksitas validasi dan kinerja sistem, yang selama ini jarang dibahas secara mendalam pada penelitian lokal sejenis. Penelitian ini memperkaya literatur tentang sistem absensi digital di Indonesia dengan menambahkan dimensi evaluasi usability dan waktu respon sebagai indikator penting dalam penilaian kinerja sistem QR Code. Selain itu, pendekatan metodologis yang menggunakan kombinasi pengujian sistem secara langsung dan evaluasi pengguna juga memberikan contoh penerapan metode prototyping yang efektif untuk pengembangan sistem berbasis web.

Implikasi penelitian ini sangat relevan bagi institusi pendidikan, organisasi, dan instansi pemerintahan yang ingin menerapkan sistem absensi digital. Hasil temuan menunjukkan bahwa penerapan QR Code dapat menghemat waktu operasional, mengurangi kesalahan pencatatan, serta meningkatkan transparansi keberadaan data. Namun, penerapan validasi tambahan harus disesuaikan dengan kondisi jaringan, perangkat pengguna, serta kapasitas server. Dalam konteks akademik, penelitian ini memberikan dasar teoritis bahwa optimalisasi kinerja sistem berbasis QR tidak hanya terletak pada keakuratan data, tetapi juga pada pengalaman pengguna (user experience) dan efisiensi waktu respon.

Penelitian ini tentu memiliki beberapa batasan. Uji coba hanya dilakukan di satu lembaga dengan jumlah pengguna terbatas dan waktu pengujian yang relatif singkat, sehingga hasilnya belum dapat digeneralisasikan ke semua konteks institusi. Selain itu, fitur validasi yang diuji masih terbatas pada selfie dan GPS, sementara teknologi validasi lain seperti *pengenalan wajah* atau biometrik belum diimplementasikan. Aspek performa yang diuji juga belum mencakup uji beban sistem (stress test) serta analisis skalabilitas server dalam kondisi penggunaan massal.

Berdasarkan keterbatasan tersebut, penelitian selanjutnya disarankan untuk memperluas cakupan uji coba ke lebih banyak institusi dengan jumlah pengguna yang lebih besar serta mengintegrasikan metode validasi berbasis biometrik atau kecerdasan buatan guna meningkatkan keamanan tanpa mengorbankan sistem kecepatan. Penelitian masa depan juga dapat mengeksplorasi penggunaan framework modern seperti Laravel atau Node.js serta penyimpanan berbasis cloud untuk meningkatkan skalabilitas. Dengan arah penelitian yang lebih luas, diharapkan sistem absensi berbasis QR Code dapat berkembang menjadi solusi yang semakin efisien, aman, dan adaptif terhadap kebutuhan digitalisasi administrasi kehadiran di Indonesia.

DAFTAR PUSTAKA

[1] M. Setiono and H. Oktafiandi, “Sistem Absensi Guru Dan Siswa Dengan Kode QR Berbasis Web (Studi Kasus SMK Muhammadiyah Purwodadi Purworejo),” *J. Ekon. Dan Tek. Inform.*, vol. 10, no. 2, pp. 1–7, 2022.

[2] Y. T. Bilqis, H. Herdianto, and H. Hendry, “Sistem Absensi Karyawan Berbasis Web Menggunakan Metode QR Code pada Kantor Desa Cinta Raja,” *J. Minfo Polgan*, vol. 14, no. 1, pp. 86–93, 2025, doi: 10.33395/jmp.v14i1.14625.

[3] Elmawati, V. Wedyawati, Nofriadiman, and M. D. R. Ardy, “Perancangan Sistem Informasi Absensi Siswa menggunakan QR Code berbasis Web pada SMK Pratama Padang,” *J. Sains dan Teknol. J.*

Keilmuan dan Apl. Teknol. Ind., vol. 24, no. 2, pp. 301–309, 2024, doi: 10.36275/zkekg643.

[4] M. Ikhsan and H. Helmina, "Design of a Web-Based Lecturer Attendance Information System Using QR Codes at Muhammadiyah University of Jambi," *Brill. Res. Artif. Intell.*, vol. 4, no. 1, pp. 94–99, 2024, doi: 10.47709/brilliance.v4i1.3838.

[5] M. Himyar, M. F. Mulya, and J. H. Siringo Ringo, "Aplikasi Absensi Karyawan Berbasis Android Dengan Penerapan QR Code Disertai Foto Diri Dan Lokasi Sebagai Validasi Studi Kasus: PT.Selindo Alpha," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 2, pp. 64–74, 2021, doi: 10.47970/siskom-kb.v4i2.186.

[6] Sayuti Malik, L. Lukman, Eneng Isti Alpiyanti, and Wasis Haryono, "Perancangan Sistem Informasi Absensi QR Code dengan Metode Prototype Pada Smp Raudlatul Hikmah," *J. Innov. Educ.*, vol. 3, no. 2, pp. 357–367, 2025, doi: 10.59841/inoved.v3i2.2958.

[7] Sutisna, F. Akbarulloh, A. A. Wahyudi, S. F. Banase, and N. I. Simarmata, "Rancang Bangun Aplikasi Absensi Karyawan menggunakan QR-Code Berbasis Web pada SMA Candra Naya," *AJAD J. Pengabdi. Kpd. Masy.*, vol. 4, no. 1, pp. 130–135, 2024, doi: 10.59431/ajad.v4i1.286.

[8] I. Nuralif and M. Fachrie, "Development of A QR Code-Based Attendance System for Factory Employees," *Int. J. Softw. Eng. Comput. Sci.*, vol. 3, no. 3, pp. 281–286, 2023, doi: 10.35870/ijsecs.v3i3.1774.

[9] N. Renaningtias and D. Apriliani, "Application of the Prototype Method in the Development of a Student Final Project Information System," *J. Rekursif*, vol. 9, no. 1, 2021, [Online]. Available: http://ejournal.unib.ac.id/index.php/rekursif/92

[10] M. Alda, M. Juarsyah, A. Nugraha, and L. R. Alfachry, "Aplikasi Absensi Mahasiswa Kerja Praktik Menggunakan QR Code Berbasis Android," *J. Manaj. Inform.*, vol. 14, no. 1, pp. 27–41, 2024, doi: 10.34010/jamika.v14i1.11775.

[11] P. Kustanto, R. Bram Khalil, and A. Noe'man, "Penerapan Metode Prototype dalam Perancangan Media Pembelajaran Interaktif," *J. Students' Res. Comput. Sci.*, vol. 5, no. 1, pp. 83–94, 2024, doi: 10.31599/6x0dfz47.

[12] A. Fadilah, M. Yahya, and E. S. Rahman, "Pengembangan Sistem Absensi Qr Code Berbasis Web Untuk Siswa Jurusan Teknik Komputer dan Jaringan SMK Negeri 4 Makassar," *JIMUJurnal Ilm. Multidisipliner*, vol. 2, no. 04, pp. 1060–1073, 2024, doi: 10.70294/jimu.v2i04.487.

[13] A. Wahyuni, "Rancang Bangun Sistem Informasi Absensi Karyawan Berbasis Website," *JIKA (Jurnal Inform.*, vol. 6, no. 1, p. 27, 2022, doi: 10.31000/jika.v6i1.5164.

[14] D. Hamdani, A. P. W. Wibowo, and H. Heryono, "Perancangan Sistem Presensi Online dengan QR Code Menggunakan Metode Prototyping Designing an Online Attendance System with QR Code Using Prototyping Method," *J. Teknol. dan Inf.*, vol. 14, no. 1, pp. 62–73, 2024, doi: 10.34010/jati.v14i1.

[15] E. Herlina and T. Hidayatulloh, "Penerapan QR Code Untuk Sistem Absensi Siswa SMP Berbasis Web," *J. Teknol. dan Inf.*, vol. 7, no. 2, pp. 102–112, 1970, doi: 10.34010/jati.v7i2.865.

[16] M. Nanda Aprillia, F. Wahyu Christanto, B. Parga Zen, H. Maulana, and N. Yudi Permana, "Implementasi Teknologi QR Code pada Sistem Absensi Karyawan Berbasis Website," *J. Sist. Inf. Galuh*, vol. 3, no. 1, pp. 39–50, 2025, doi: 10.25157/jsig.v3i1.4445.

[17] H. dan Irwandi, "sistem absensi mahasiswa menggunakan QRCODE berbasis mobile dan website pada universitas fajar," pp. 1–23, 2022.

e-ISSN3032-601X & p-ISSN3032-7105

[18] S. W. Taju, Y. P. Mamahit, and J. A. Pongantung, "Implementing QR code and Geolocation Technologies for the Student Attendance System," *CogITo Smart J.*, vol. 10, no. 1, pp. 221–232, 2024, doi: 10.31154/cogito.v10i1.636.642-653.

[19] P. Metra, D. Doni, D. Sodikin, and M. Azikri, "Implementasi Sistem Absensi Online Berbasis Foto Selfie dan Deteksi Lokasi di Dinas Kehutanan Provinsi Jambi," *RIGGS J. Artif. Intell. Digit. Bus.*, vol. 4, no. 2, pp. 5001–5008, 2025, doi: 10.31004/riggs.v4i2.1380.